From "Anthropomorphic Attribution"to "Alliance Formation" : The Impact of Human-Chatbot Relationship on Engagement

Authors: Mo Ran, Fang Jiandong, Chang Baorui, Chang Baorui

Date: 2023-04-03T00:00:00+00:00

Abstract

With the rapid advancement of Artificial Intelligence (AI) technology, AI chatbots can simulate human guidance to improve user engagement and therapeutic efficacy in Internet-based Self-help Interventions (ISIs). However, academic exploration of chatbot mechanisms of action remains in its infancy. Therefore, to deepen theoretical understanding of this issue, this article proposes a theoretical model adapted to the ISI context from a human-computer relationship perspective: chatbots can progressively develop Human–Chatbot Relationships (HCRs) with users through four stages—anthropomorphic attribution, utilitarian value judgment, developing attachment relationships, and establishing a Digital Therapeutic Alliance (DTA)—and thereby enhance user engagement through HCRs. Future research may continue to enrich HCRs-related theories and examine their underlying mechanisms, design chatbots based on HCRs theory, thoroughly investigate additional variables influencing HCRs, unify the operational definition of engagement, and develop appropriate measurement tools for engagement.

Full Text

From Anthropomorphic Attribution to Alliance Establishment: The Effect of Human-Chatbot Relationships on Engagement

Mo Ran[1], **Fang Jiandong**[1,2,3]#, **Chang Baorui**[1,2,3]

[1]Department of Psychology, Faculty of Education, Guangxi Normal University, Guilin 541006, China
[2]Guangxi University and College Key Laboratory of Cognitive Neuroscience and Applied Psychology, Guangxi Normal University, Guilin 541006, China

[3]Guangxi Ethnic Education Development Research Center, Key Research Base of Humanities and Social Sciences in Guangxi Universities, Guilin 541006, China

Abstract: With the rapid development of Artificial Intelligence (AI) technology, AI chatbots can simulate human guidance to improve user engagement and efficacy in Internet-based Self-help Interventions (ISIs). However, research on the mechanisms underlying chatbot effectiveness remains in its early stages. To deepen our theoretical understanding of this issue, this article proposes a theoretical model adapted to the ISI context from a human–computer relationship perspective: chatbots can develop Human–Chatbot Relationships (HCRs) with users through four sequential stages—anthropomorphic attribution, utilitarian value judgment, attachment relationship development, and establishment of Digital Therapeutic Alliance (DTA)—thereby enhancing user engagement through HCRs. Future research should continue to enrich HCRs theory and examine its underlying mechanisms, design chatbots based on HCRs theory, investigate additional variables affecting HCRs, unify operational definitions of engagement, and develop appropriate engagement measurement tools.

Keywords: chatbot, engagement, human–chatbot relationships

With the rapid development of mobile internet, Internet-based Self-help Interventions (ISIs) have emerged as more flexible, economical, and convenient alternatives to traditional psychological counseling and psychotherapy (Mrazek et al., 2019). In recent years, substantial evidence has supported the feasibility and effectiveness of ISIs (Izzaty et al., 2021; Johansson et al., 2021; Sun et al., 2021; Taylor et al., 2021; Weisel et al., 2019). However, high dropout rates and low user engagement remain significant challenges (Taylor et al., 2021). Engagement refers to the extent to which users experience or participate in ISI content (Christensen et al., 2009; De Geest & Sabaté, 2003; Leeuwerik et al., 2019) and is closely linked to treatment efficacy (Asaeikheybari et al., 2021; Cavanagh et al., 2018; Karyotaki et al., 2017; Puls et al., 2020; Tetley et al., 2011). While human support can enhance engagement (Baumeister et al., 2014; Richards & Richardson, 2012; Spijkerman et al., 2016), such support limits the immediacy, scalability, and cost-effectiveness of ISIs. Fortunately, AI-powered chatbots embedded in ISIs can simulate human guidance and have been shown to promote greater user engagement and efficacy compared to traditional ISI programs (Perski et al., 2019; Provoost et al., 2020b; Vaidyam et al., 2019). Nevertheless, research in this domain, particularly outside China, remains focused on examining feasibility and effectiveness, with limited understanding of the underlying mechanisms (Bassi et al., 2022; Gabrielli et al., 2021; He et al., 2022; Skjuve et al., 2022b). Therefore, deepening our theoretical understanding is essential for designing more targeted chatbots that further enhance ISI outcomes. This article explains the mechanisms of ISI chatbots through the lens of Human–Chatbot Relationships (HCRs), discussing the developmental process of HCRs in ISI contexts and its impact on user engagement to inform future research.

The Developmental Process of HCRs and Its Impact on User Engagement

From a Human–Computer Interaction (HCI) perspective, low engagement stems from inadequate interactive experiences, making the static interactions of traditional ISI programs ill-suited for fostering user participation (Doherty et al., 2012). In contrast, chatbots with natural language conversation capabilities enable active user participation rather than passive information reception. These agents range from simple chat avatars to virtual embodiments and function as more active social actors (Appel et al., 2012; Doherty et al., 2012; Go & Sundar, 2019; Ly et al., 2017). Moreover, conversational behaviors—both verbal and nonverbal—are necessary conditions for relationship formation (Bickmore & Picard, 2005), suggesting that chatbots can establish and develop HCRs with users, thereby enhancing engagement through relational bonds.

Given the similarities between HCRs and Human–Human Relationships (HHRs), several HHRs theories can help explain HCRs development (Hendriks et al., 2020; Schuetzler et al., 2020). For instance, Skjuve et al. (2021a) developed a three-stage model to understand HCRs development based on the well-validated Social Penetration Theory (SPT): Stage 1 (Exploration) involves cautious user attitudes toward chatbots due to privacy or safety concerns, characterized by superficial self-disclosure; Stage 2 (Affection) involves utilitarian value judgment and attachment development, promoting interaction frequency and disclosure depth; and Stage 3 (Stability) involves maintained attachment with chatbots becoming part of daily life, though with reduced self-disclosure and increased sharing of routine events. Subsequently, Skjuve et al. (2022b) validated this model through a 12-week longitudinal study, revealing the progressive nature of HCRs development. However, despite its novelty, this model has several limitations when applied to ISIs: it neglects cognitive processing at the initial stage of human–computer interaction, fails to explain psychological mechanisms underlying each HCRs stage, and focuses primarily on intimate relationship development without considering the specificities of psychological counseling and psychotherapy contexts. Therefore, this article integrates HCI and psychological theories to refine and contextualize the HCRs model for digital mental health applications.

2.1 Stage 1: Anthropomorphic Attribution

Anthropomorphism is the cognitive process through which humans attribute human-like features (e.g., appearance, speech), motivations, intentions, or emotions to non-human entities (e.g., inanimate objects, animals) (Epley et al., 2007), characterized by heuristic processing (Tversky & Kahneman, 1974). Generally, when users first encounter a chatbot, anthropomorphism is immediately activated. Users unconsciously treat it as another person (hereafter "anthropomorphic attribution") based on its human-like appearance or conversational ability, leading them to adopt interpersonal interaction strategies (Nass et al., 1994). With increased interaction frequency, users may progressively attribute

deeper characteristics such as motivations, intentions, and emotions (Xu et al., 2017), and strengthened anthropomorphism further promotes HCRs development (Pentina et al., 2023). Thus, anthropomorphic attribution may be a critical starting point for HCRs development, and understanding its underlying mechanisms can facilitate this process.

First, users engage in anthropomorphic attribution driven by intrinsic motivation. Based on Reeves and Nass' s (1996) Media Equation Theory (TME), "media equals real life," meaning people exhibit social, instinctive reactions to media even when aware of their irrationality. The Computers Are Social Actors (CASA) paradigm, a key research framework within TME, posits that users intuitively attend to human cues presented by computers (e.g., text output, voice, language style, interactivity) while ignoring instrumental cues, naturally perceiving them as social actors and responding with social rules (e.g., bias, politeness, reciprocity) that generate para-social reactions (e.g., trust, liking) (Nass et al., 1994). But why do users make such unconscious reactions? According to Epley et al.' s (2007) three-factor theory, anthropomorphism comprises three synergistic factors: Elicited Agent Knowledge, Effectance Motivation, and Sociality Motivation. Therefore, at the initial stage of human–computer interaction, users focus on primary cues (appearance, speech) for anthropomorphic inference and attribution, providing a cognitive foundation for satisfying effectance and sociality motivations.

Second, the specific context of ISIs promotes positive user perceptions. Based on the Reduced Social Cues (RSC) theory, computer-mediated communication reduces social cues due to bandwidth limitations, and these limited cues are amplified through compensatory effects, making users'psychological states more susceptible to influence (Tanis & Postmes, 2003) and facilitating perceptual fluency —a subjective experience of pleasure and relaxation (Labroo et al., 2008). The well-validated Hyperpersonal Interaction (HI) theory similarly suggests that reduced social cues in human–computer interaction lead users to idealize chatbots, overlook technical flaws, and employ more impression management strategies to enhance "liking" (Walther, 1996).

Finally, users' social presence is enhanced, influencing their engagement. Social presence refers to the feeling of being with another person, encompassing copresence, psychological involvement, and behavioral engagement (Biocca et al., 2003). Influenced by intrinsic motivation and limited social cues, users' "social presence heuristic" is effectively activated (Sundar et al., 2008), representing a core element through which chatbots influence engagement at the initial stage of HCRs development. Research indicates that when social presence is activated by chatbots, users' behavioral intention significantly increases (Mozafari et al., 2021). Additionally, a meta-analysis by Blut et al. (2021) found that social presence mediates the effect of anthropomorphism on behavioral intention. According to the Theory of Planned Behavior (TPB), behavioral intention is a key factor influencing users' intrinsic motivation and predicting actual participation behavior (Ajzen, 2012). Furthermore, if users perceive interactions with

chatbots as vivid as those with real people, trust is more easily established, and engagement consequently improves (Brendel et al., 2022; Hassanein & Head, 2005; Lee et al., 2021).

In summary, effective anthropomorphic attribution (primary cues like appearance and speech) is a prerequisite for HCRs development. Users engage in parasocial interaction with chatbots as they would with humans, providing a cognitive foundation for satisfying effectance and sociality motivations. As HCRs develop, anthropomorphism gradually becomes primarily motivation-driven with attribution as a secondary process, aiming to better satisfy individual needs. Driven by effectance and sociality motivations, users amplify limited social cues, thereby enhancing social presence and influencing engagement, while anthropomorphic attribution deepens (advanced cues like motivations, intentions, emotions), further promoting HCRs development.

2.2 Stage 2: Utilitarian Value Judgment

Utilitarian value refers to the degree to which a product satisfies users' functional needs, such as information acquisition, efficiency improvement, and problem-solving (Choi & Drumwright, 2021). In the early stages of HCRs development, users tend to judge chatbots' utilitarian value to determine whether their actual needs can be met. Specifically, whether chatbots can demonstrate practical utility according to users' current expectations—such as accurate and personalized mental health information, smooth and natural conversational ability, precise and intelligent context comprehension, and rich, high-quality skill services—affects user acceptance. Therefore, examining the mechanisms of utilitarian value judgment can help specify chatbot functional design and promote HCRs development.

First, users tend to position chatbots as "tools" and focus on their practicality. On one hand, while anthropomorphic attribution facilitates positive initial impressions, users' trust has not yet been established, preventing deeper self-disclosure and resulting in superficial HCRs at this stage (Skjuve et al., 2021a). Additionally, research indicates that humans hold stereotypes about chatbots, acknowledging their ability to challenge human intelligence but believing they fundamentally lack emotional capacity. This perception stems from humans' tendency to interpret chatbot responses as programmed computations rather than "spontaneous" reactions, hindering deeper emotional interaction (Wirtz et al., 2018). On the other hand, based on the Uses and Gratifications Framework (Rubin, 1983), users are active, rational, and goal-oriented, seeking and selecting media products that satisfy specific needs (e.g., interaction, functionality, entertainment, information acquisition, social status). This explains why users prioritize chatbots' practical utility during initial HCRs development.

Second, utilitarian value judgment influences users' subjective attitudes. According to the well-validated Technology Acceptance Model (TAM), users' functional needs for technology can be summarized as perceived usefulness and perceived

ease of use, which determine attitudes toward utilitarian value and influence behavioral intention (Legris et al., 2003). For example, in Kamita et al.' s (2019) ISI study, chatbots scored significantly higher than traditional web programs on both usefulness and ease of use, and chatbot group participants (N=15) showed greater engagement and significantly higher behavioral intention scores than web program participants (N=12). Similarly, Park and Kim (2023) found that perceived usefulness positively predicted willingness to socially interact with mental health chatbots. The Expectation-Confirmation Model (ECM), an extension of TAM, proposes a mechanism for evaluating utilitarian value: users compare pre-use expectations with post-use perceived usefulness to determine whether expectations are confirmed, which determines satisfaction (Bhattacherjee, 2001). For instance, Dhiman and Jamwal (2022) used ECM to investigate continued chatbot use, finding that perceived usefulness and post-use expectation confirmation significantly affected satisfaction. Xie et al. (2022) found utilitarian value to be the strongest predictor of user satisfaction compared to technological, hedonic, and social factors.

Finally, users' subsequent engagement behavior is influenced by both behavioral intention and satisfaction. Behavioral intention, defined as the strength of willingness to perform an action, directly affects engagement motivation and actual participation behavior (Ajzen, 2012). High behavioral intention leads to interaction with chatbots and participation in ISI programs, whereas unstable motivation makes subsequent engagement unlikely (Alfonsson et al., 2017). User satisfaction, as a more comprehensive evaluation indicator, affects both engagement motivation and loyalty, representing an important factor for continued interaction (Cheng & Jiang, 2020). For example, Zhu et al. (2022) found that chatbots' utilitarian value (personalized information presentation) significantly improved user satisfaction, which was positively correlated with continued use intention. However, in Liu et al.' s (2022) 16-week ISI study, chatbot group engagement declined over time, with researchers attributing this to both technical deficiencies and failure to deliver useful, satisfying content.

Overall, this stage represents users' exploration phase, during which they may hold stereotypes of chatbots as emotionless tools. Based on the Uses and Gratifications Framework, chatbots must first demonstrate value as effective tools by proving their importance in terms of usability, ease of use, and expectation confirmation to promote engagement. Applying Epley et al.' s (2007) three-factor theory, as interaction frequency increases, users' recognition, familiarity, and sense of certainty about chatbots improve. This process satisfies users' need to understand, predict, and control unfamiliar objects (effectance motivation), thereby strengthening anthropomorphism and indirectly enhancing liking, which further develops HCRs.

2.3 Stage 3: Developing Attachment Relationships

Attachment refers to the enduring and stable emotional bond that individuals develop with significant others during infancy, which influences their attribution

patterns and gradually internalizes into unique attachment styles that serve as templates for future relationships with friends, family, romantic partners, or possessions (Bartholomew & Horowitz, 1991). The key factor influencing attribution is the level of security experienced; each interpersonal interaction involves applying past attachment styles, evaluating security obtained, and making relationship-developing behaviors (approach or avoidance) (Adams et al., 2018). Recently, attachment theory has been widely applied in communication research to explain emotional attachment development to non-human objects (e.g., pets, brands, virtual characters) and motivations for relationship maintenance (Bauer & Woodward, 2007; Pedeliento et al., 2016; Wanser et al., 2019; Xie & Pentina, 2022). As HCRs further develop, chatbots as active social actors may also develop attachment relationships with users.

First, human cognitive and affective processing mechanisms facilitate human-chatbot attachment. Research shows that when users continuously interact with chatbots exhibiting human-like features, particularly relational cues (e.g., humor and empathy), their cognitive reasoning drives emotional perception, which in turn influences cognitive judgments (Lee, S. et al., 2020; Sánchez-Franco et al., 2021; Spatola & Wudarczyk, 2021). In Beck' s (1995) cognitive model, cognition and emotion mutually influence each other. Furthermore, based on Mind Perception Theory, users integrate cognitive and affective dimensions to process anthropomorphic information about chatbots, forming perceived mind attributions that influence interaction willingness (Waytz et al., 2010; Blut et al., 2021). The well-known Dual-Process Theory also indicates that cognitive processes comprise rational components (e.g., utilitarian value perception) and irrational emotional components (e.g., emotional value perception) (Stanovich & West, 2000). Thus, cognition and emotion complement each other in human-computer interaction, shaping user experience and fostering attachment relationships between chatbots and users (Abdulrahman & Richards, 2021; Bickmore et al., 2005; Choi & Drumwright, 2021; Pentina et al., 2023; ter Stal et al., 2020).

Second, chatbots can possess key conditions for developing attachment relationships. On one hand, chatbots are more reliable and controllable while serving "safe base" and "haven" functions, generating attachment by meeting spiritual needs or alleviating distress (Rabb et al., 2022). For example, Zhou et al. (2020) described how the chatbot "Xiaoice" successfully established attachment relationships with users through its unique emotional intelligence system that satisfied multiple spiritual needs (communication, emotion, social belonging), promoting long-term sustained activity. Xie and Pentina (2022) similarly found that when participants perceived understanding and received appropriate emotional support and encouragement from the chatbot Replika during times of distress and loneliness, they developed attachment. On the other hand, chatbots' natural language conversation advantage enables their "disclosure" to promote interaction frequency and user self-disclosure, allowing users to experience acceptance, intimacy, and loneliness relief as continuous intrinsic rewards that develop attachment (Skjuve et al., 2021a, 2022b). In Social Penetration Theory, increased information transfer and gradual self-disclosure represent trust, which

is considered a key prerequisite for relationship development (Altman & Taylor, 1973). For instance, Kang and Gratch (2014) found that participants interacting with deeply self-disclosing chatbots responded with greater self-disclosure and reported higher trust and intimacy. Lee, Y. C. et al. (2020) similarly found that deeply self-disclosing chatbots promoted more user self-disclosure than shallow or non-disclosing chatbots and successfully established attachment relationships with some participants.

Finally, attachment relationship development promotes longer-term ISI participation. While anthropomorphic attribution and utilitarian value judgment provide cognitive foundations and promote social presence and interaction frequency, these represent short-term perspectives. Focusing solely on cognitive aspects results in shallow anthropomorphism (e.g., external appearance, speech human-likeness). Moreover, more efficient and practical chatbots become positioned as "tools" rather than feeling, trustworthy "partners," weakening user connection and reducing engagement. In Mind Perception Theory, individuals deny "humanness" to human-like objects perceived as lacking emotional capacity and refuse equal communication (Waytz et al., 2010). Conversely, social and emotional needs for contact, connection, and recognition are fundamental relationship demands and important factors for long-term, stable ISI participation (Epley et al., 2007). Therefore, to strengthen HCRs stably, users must actively participate in human–chatbot interaction, and chatbots must "act" to provide "emotional value." Driven by social motivation, users may further anthropomorphize chatbots by attributing deeper characteristics (motivations, intentions, emotions) (Xu et al., 2017) and establish attachment relationships to satisfy emotional needs more continuously and fully (Xie & Pentina, 2022; Zhao et al., 2012). HCRs development thus deepens through strengthened anthropomorphism (Epley et al., 2007; Pentina et al., 2023), further improving engagement.

In summary, anthropomorphic attribution and utilitarian value judgment are short-term factors affecting ISI engagement. To maintain user activity over longer periods, emotional factors play a more important role. Driven by social motivation, users further anthropomorphize chatbots and establish attachment bonds, experiencing more positive emotions while HCRs deepen—from "tool" to "partner." If users transfer attachment from chatbots to ISI tasks, they become more likely to participate actively and continuously (McGonagle et al., 2021), thereby achieving the goal of using chatbots for psychological counseling and therapy.

2.4 Stage 4: Establishing Digital Therapeutic Alliance

Therapeutic Alliance (TA) refers to the collaborative relationship between client and counselor in therapy (Zhu & Jiang, 2011) and is a robust predictor of treatment outcomes (Flückiger et al., 2018). Research shows that TA is not limited to human–human relationships; humans can unconsciously form alliances with virtual programs. This alliance in ISI contexts is termed Digital Thera-

peutic Alliance (DTA)—the "collaborative" relationship between humans and programs (Berger, 2017; D' Alfonso et al., 2020; Darcy et al., 2021; Heim et al., 2018). Studies have found that DTA shares conceptual invariance with TA and can predict improved efficacy (Luo et al., 2022). Additionally, TA' s emotional bond dimension resembles secure attachment and correlates highly with it, while insecure attachment—particularly avoidant attachment—is key to TA rupture (Mallinckrodt & Jeong, 2015; McGonagle et al., 2021). In ISI contexts, Hertlein and Twist (2018) found that avoidant users may exhibit low activity and engagement (rejection), while anxious users may overuse programs (dependence). Therefore, understanding HCRs development through the DTA perspective is more appropriate for ISI contexts.

First, establishing and developing DTA better facilitates ISI goal achievement. In actual psychological counseling, dual relationships should be avoided, and this principle applies to ISIs as well. The purpose of establishing DTA is not to increase user acceptance or satisfaction with chatbots but to better motivate positive change. DTA makes ISI processes more transparent, requiring chatbots to reach agreements with users on goal achievement and implementation while being open about technologies and privacy policies (Law et al., 2022). Thus, DTA represents a deliberate and purposeful human–computer relationship model. Given that ISIs aim to improve users' psychological problems, this stage should regulate HCRs development toward collaborative alliance rather than other attachment relationships (e.g., friends, partners).

Second, building on the first three stages, establishing DTA becomes easier. Based on previous TA definitions—whether the classic three-dimensional structure (task agreement, goal agreement, emotional bond) (Bordin, 1979) or four-dimensional structure (openness, trust, collaborative relationship, emotional bond) (Agnew-Davies et al., 1998)—both contain cognitive and affective components: cognitive components involve recognition of therapeutic goals and tasks, while affective components involve positive emotional connections or personal attachment (e.g., mutual trust, liking, respect, care, frankness) (Zhu & Jiang, 2011). In ISIs, cognition and emotion are similarly considered key pathways to promoting DTA (D' Alfonso et al., 2020; Tong et al., 2022). On one hand, following anthropomorphic attribution, users unconsciously treat chatbots as social actors and, through rational cognitive processes (utilitarian value judgment), form cognitive evaluations of whether chatbots meet their needs, facilitating "consensus" formation. On the other hand, as HCRs develop, users develop emotional attachment, influencing emotional bond establishment. Thus, the staged development of HCRs lays a solid foundation for DTA.

Finally, promoting DTA development further enhances engagement. Regarding DTA promotion, when individuals interact with a caring object during times of need, their sense of security strengthens, and they experience feeling loved and cared for. If this influence is repeated consistently over time, insecure attachment may gradually transform into secure attachment (Bowlby, 1988; Mikulincer & Shaver, 2020; Nanjappa et al., 2014). Therefore, interactions

that enhance security can activate secure attachment perception, strengthening emotional bonds and promoting DTA development. For specific interaction design, strategies used by counselors to develop TA can inform the design of relational cues for chatbots (Mo et al., 2023). In traditional counseling, relationships develop rapidly and healthily when counselors are empathetic, genuine, and provide unconditional positive regard (Rogers, 1957). Similarly, Skjuve et al. (2021a) argue that the key for chatbots to establish and develop DTA is providing emotional support through relational cues such as acceptance, understanding, and non-judgment. Bell et al. (2019) also emphasize that lacking understanding and empathy prevents chatbots from establishing the relationship strength needed for effective psychotherapy. In short, chatbots can first simulate counselor identity to create a professional, reliable initial impression, then rely on good functional experience to gain user recognition and increase reuse rates, and further present relational cues that strengthen emotional bonds (e.g., friendliness, respect, non-judgment, listening, encouragement, genuineness, empathy, trust, self-disclosure) while adhering to ethical guidelines (Mo et al., 2023). This approach can regulate attachment relationships into more collaborative alliance relationships conducive to ISI goals during HCRs development. Once DTA is established, users exhibit reduced self-protection, deeper self-disclosure, and greater willingness to collaborate, leading to improved engagement and treatment outcomes (Heim et al., 2018; Liu et al., 2022; Provoost, 2021a). For example, Goldberg et al. (2021) found that a fully automated mindfulness intervention program could establish DTA with participants, which not only significantly predicted engagement but also predicted depression and anxiety improvement at weeks 3 and 4. In Rodrigues et al.' s (2021) randomized controlled trial, participants similarly established DTA with a chatbot, which positively predicted engagement.

In summary, HCRs development may significantly impact user engagement in ISIs. High engagement requires users to remain active throughout longer ISI lifecycles, presenting a major challenge for the field. This article integrates mature theories from HCI and psychology to refine the HCRs development framework and proposes a theoretical model adapted to ISI contexts (Figure 1 [Figure 1: see original paper]).

Task/Goal Agreement → Media Equation Theory → Three-Factor Theory → Reduced Social Cues → Utilitarian Value → Uses and Gratifications → Technology Acceptance Model → Expectation-Confirmation Model → HCRs Promoting Engagement Theoretical Model (Note: Arrows indicate possible causal directions)

Future Research Directions

3.1 HCRs Theory Is Underdeveloped and Mechanisms Remain Unclear

Although Bickmore and Picard (2005) provided important insights into how relational cues (e.g., humor and empathy) affect HCRs development, they did not construct a theory to explain HCRs changes. To date, no consensus exists on HCRs development theory, and researchers still know little about how HCRs are initiated, develop, strengthen, and affect humans, indicating substantial room for new theory development (Muresan & Pohl, 2019; Skjuve et al., 2021a, 2022b). Since HCRs development may parallel HHRs development, existing HHRs theories can serve as starting points. Future research could draw on mature HHRs theories such as Social Exchange Theory (Emerson, 1976), the Investment Model of Personal Relationships (Rusbult et al., 1994), and Commitment-Trust Theory (Morgan & Hunt, 1994) to deepen understanding of HCRs development. Moreover, the mechanisms through which HCRs stages affect human psychology are complex, and the importance of human cues may vary across stages. For example, visual and verbal cues significantly impact satisfaction in early HCRs development (Kim et al., 2021), while nonverbal and relational cues may become more important in middle and later stages. Future research should employ Digital Psychometrics (Latynov & Shepeleva, 2020) and longer-term longitudinal studies to rigorously test different HCRs stages and derive effective design principles.

3.2 Insufficient Consideration of Chatbot Cues

Most current ISI studies use highly task-oriented chatbots with inadequate consideration of human cues. Researchers often only examine whether chatbots affect dependent variables while ignoring the importance of human cues, leading to poor comparability and reproducibility (Chong et al., 2021). Future studies should report which human cues are implemented and discuss their relationships with dependent variables. Additionally, only a few studies have examined relationships between human cues and outcomes, yet these cue designs lack theoretical grounding (Rapp et al., 2021). Although HHRs theories are important for understanding HCRs development, researchers still know little about which factors promoting HHRs development are effective and important in HCRs, and how HHRs and HCRs development differ. Future research should design human cues based on theory and test their contributions through rigorous randomized controlled trials. Furthermore, digital mental health encompasses multiple scenarios (e.g., intelligent triage, emotional companionship, counseling, psychotherapy) requiring different HCRs levels. For instance, intelligent triage involves short user lifecycles with higher efficiency demands, making anthropomorphic attribution and utilitarian value judgment more important. Researchers should evaluate application scenarios and design specialized human cues accordingly. Finally, users from different cultural backgrounds, genders, age groups, personality traits, education levels, income levels, and symptom

profiles may perceive different chatbot identities and corresponding human cues differently (Nißen et al., 2022). Future research should examine optimal identity and human cue combinations for different user characteristics and scenarios.

3.3 No Standard for Engagement Measurement and Reporting

Although preliminary evidence supports the feasibility and effectiveness of chatbots for psychological intervention, these studies lack consensus on engagement evaluation standards. First, researchers often conflate "adherence" with "engagement" (Beintner et al., 2019; Eysenbach et al., 2011). However, "adherence" is more commonly applied clinically, emphasizing a doctor–patient relationship. Since most ISI users are not patients, this article uses "engagement" as an umbrella term describing ISI usage, implying an equal client–counselor relationship requiring active user participation. Second, engagement reporting methods vary considerably, with most studies failing to report engagement–treatment outcome relationships (Beintner et al., 2019; Vaidyam et al., 2019). Ignoring this relationship in ISI research may lead to underestimated intervention effects and reduce comparability across studies. Third, most ISI studies use single, theoretically unsupported engagement indicators (e.g., "number of completed exercises"), though numerous indicators are available (Lederman & D' Alfonso, 2019). Future research should unify and enrich engagement evaluation indicators based on theory (Beintner et al., 2019). Finally, most ISI studies over-rely on self-report methods, which may overestimate engagement (Flett et al., 2019). Future research should integrate objective data (program backend data, wearable device biosignals, demographic variables) with subjective self-report data for a more comprehensive understanding of engagement.

3.4 Additional Variables Affecting HCRs Require Investigation

Beyond chatbot human cues, numerous factors in actual ISIs can confound results. First, product performance: chatbots exhibit poor stability in open domains, with frequent technical errors (Jang et al., 2021) and repetitive, unnatural conversational experiences (Fulmer et al., 2018), hindering HCRs development. Additionally, research suggests that ISI programs as a whole are more important than individual functions (Berger et al., 2014), leaving unclear whether chatbots or the overall ISI produce key improvements. Second, privacy: developing ISI programs risks violating large amounts of private data (McGreevey et al., 2020). If users perceive their stored data as insecure, engagement likely decreases, potentially leading to dropout (Proudfoot et al., 2010). Third, novelty effect: curiosity about new technologies may promote short-term enthusiasm, overestimating engagement (Croes & Antheunis, 2021; Fryer et al., 2017; Nadarzynski et al., 2019). Fourth, uncanny valley effect: according to the Theory of the Uncanny Valley, human-like objects that become too realistic may evoke dislike (Mori et al., 2012; Song & Shin, 2022). Given rapid advances in AI language models and chatbots' increasingly sophisticated simulation capabilities (e.g., ChatGPT) (Aydın & Karaarslan, 2022; Elkins & Chun, 2020),

future HCRs research should consider this effect. Fifth, interactions among human cues: different chatbot identities (robot vs. human) affect user expectations and consequently influence effects of other human cues (e.g., appearance, natural language ability), suggesting possible cue interactions where different combinations affect outcomes differently (Go & Sundar, 2019). In summary, future research should objectively evaluate chatbot effects while controlling for these additional variables to improve result reliability.

References

Mo, R., Fang, J. Z., & Fang, J. D. (2023). How to establish a digital therapeutic alliance between chatbots and users: The role of relational cues. *Advances in Psychological Science*, 31(4), 669-683.

Xu, L. Y., Yu, F., Wu, J. H., Han, T. T., & Zhao, L. (2017). Anthropomorphism: From "it" to "he" . *Advances in Psychological Science*, 25(11), 1941-1953.

Zhu, X., & Jiang, G. R. (2011). The concept of working alliance. *Chinese Journal of Clinical Psychology*, 19(2), 275-280.

Zhao, X., Zhou, M., Yu, L. L., & Liu, Q. (2012). A model of virtual community continuance from an emotional attachment perspective: Beyond the cognitive judgment paradigm. *Frontiers of Engineering Management*, 31(5), 14-20.

[The remaining references are preserved exactly as provided in the original text.]

Note: Figure translations are in progress. See original paper for figures.